

Aggression and aggressiveness in social media: research on Russian-language data

Soloviev A.N. (solovyev@br-analytics.ru)

Brand-Analytics, Moscow, Russia

Annotation. In this paper, from a linguistic point of view, we consider such topical concepts as aggressiveness and aggression. In particular, aggression is defined as one of the functions of aggressiveness. The types of aggression are proposed not as invective categories (trolling, racism, cyberbullying etc.), but as functional types of aggressive discourse. The results of the study of the aggressiveness in the texts and its dynamics are demonstrated as a case study of social media.

Keywords: aggression, aggressiveness, functions of aggressiveness, types of aggression, aggressiveness in social media.

Агрессия и агрессивность: исследование на материалах русскоязычных социальных медиа

Соловьев А.Н. (solovyev@br-analytics.ru)

Brand-Analytics, Москва, Россия

Аннотация. В данной работе с лингвистической точки зрения рассмотрены такие актуальные понятия как агрессивность и агрессия. В частности, агрессия определена как одна функций агрессивности. Типы агрессии представлены не как инвективные категории, а как функциональные типы агрессивного словоупотребления. Продемонстрированы результаты исследования функционирования агрессивности в текстах и ее динамики на материалах соц.медиа.

Ключевые слова: агрессия, агрессивность, функции агрессивности, типы агрессии, агрессивность в соц.медиа.

1 Introduction

The main difference between artificial intelligence (AI) and human intelligence is that the computer has no emotions. The modern concept of AI can rather be compared with elements of the cerebral cortex: fuzzy logic, classification, clustering. The limbic and paralimbic brain systems have not yet fallen into the scope of AI modeling. Nevertheless, a computer can be trained to recognize not only the basic emotional units (affectives), but also some implicit components (the so-called connotatives and potentials) [13] in a text. But if you take the point of view that consciousness is encoded by emotions which determine the thinking process [4; 5], then it is too early to talk about a genuine AI.

Aggressiveness is an essential feature of every person, the driving force of evolution, one of the fundamental emotions of all living beings [12]. In ethology, psychology and sociobiology there is a huge number of scientific researches, directions and schools. The research of aggressiveness from the point of view of its reflection in a natural language is still rather poorly represented in the scientific and technical sphere. And basically, these works deal mainly not so much with language but with speech, where more attention is paid to the psycho-physiological characteristics of speech (see the review [11]).

However, technological development needs new security requirements in cyberspace. Modern linguistic algorithms make it possible to identify not only potential threats, but also to determine the type and direction of aggression, to identify the source and its cause. So potentially important components of aggression are harassment, racism, sexism, trolling, cyberbullying, and some others (see links [1, 3]).

There are many types of aggression classifications. Type of classification depends on the goals and tools of the researcher. For example, in [2, 10] the authors define two types of aggression: 1) explicit, directed at a subject, and 2) implicit, hidden form of aggression. There are also active and passive aggression, direct and indirect, directed at a person or group of people. We will be interested in lexical-semantic aspects of aggressiveness, expressed explicitly. In result of this, the type of classification we have chosen will be based on the emotive characteristics of the word, obviously marking of aggression.

Our research goal was not just to categorize aggression by social significant issues, as in the mainstream studies in this field, but to study the functional manifestations of aggressiveness in texts and to demonstrate its dynamics as a case study of social media.

But what is aggressiveness and aggression? Let us define the basic concepts.

1.1 Aggressiveness and Aggression

Each field of knowledge determines the aggression and aggressiveness depending on the tasks. Sometimes researchers do not distinguish between them, which results in confusion. In this work, we will introduce terminology based on some works in sociobiology and ethology [12].

We will talk about the manifestation of aggressiveness and aggression by means of a natural language. Therefore, the terms linguistic aggressiveness and linguistic aggression were introduced. In contrast to verbal aggression, we introduce a unifying term, since the subject of study is rather texts than speech.

Despite the variety of definitions of aggression, one can note the main, fundamental criterion in all of them: **aggression** is understood as a directional impact on an object with the aim of causing damage. In general, this definition satisfies our objectives for the research of linguistic aggression, if we clarify what damage is and how it can be done using text. First of all, this is a deliberate reduction of the status of the subject or object of the message, which can be expressed both as a direct impact on the subject and indirectly as a description of one's expressive-negative attitude towards something or someone. It should be noted that linguistic aggression, in contrast to the biological and ethological formulations, can be not only the fact of such an impact, but also an aggressive invention («я убью тебя!») and an active call for aggressive actions («да пошли ты его!»).. Perhaps the main necessary but not sufficient condition for the manifestation of linguistic aggression is violation of language behavior norms. Sufficient conditions should include orientation of aggressive vocabulary (both explicit and implicit) towards the object or subject.

Aggressiveness, in contrast to aggression, is accumulated negative energy, which can potentially become aggression, and can ritualize (for example, in the form of crying, laughing, etc.).

From the linguistic point of view, the violation of language behavior norms is inherent for aggressiveness. But, unlike aggression, the vocabulary used is not intended to an impact or call for impact in order to cause damage, but potentially contains prerequisites for the emergence of aggression. Thus, aggressiveness, in contrast to aggression, is not an effect on an object or a subject, but a language reaction of a subject due to its internal dissatisfaction by, for example, the external environment's negative impact on it.

In other words, aggression can be viewed as an emotion (and a generative emotion generating others), and aggressiveness as a type of behavior or deviation from the norm of behavior.

1.2 Functions of linguistic aggressiveness

Let us define the functions of linguistic aggressiveness. Since we are talking about linguistic aggressiveness, its functions will partly coincide with the functions of injunctive lexicon. Well-known modern linguist Steven Pinker identifies five such functions [7]:

1. Descriptive (the meaning of the word corresponds with its subject of reference);
2. Ideomatic (or allegorical);
3. Aggressive (expressing one's assessment of someone or something); in fact, aggression itself;
4. Emphatic (roughening of semantics in order to emphasize something significant to the speaker)
5. Cathartic - a poorly controlled violent throw-out of negative emotions due to an external negative impact (for example, hitting a finger with a hammer).

All of these functions, with the exception of descriptive, are more related to aggressiveness. The descriptive function is poorly related to the aggressiveness functions, because it is characteristic of people whose speech behavior is based on invective vocabulary, the use of which is the norm for them.

Aggressive function can be disassembled into its components:

- directed action on the subject (type A) or object (type B);
- assessment of what is happening - type C;
- aggressive intentions or call for aggressive actions (type D).

The cathartic function of aggressiveness refers to pure emotion, all other types to behavior. It is not always possible to separate these types of aggression. Nevertheless, as will be shown later, aggressive function as a type of behavior prevails in its diversity, but inferior in frequency of use of the cathartic function.

2 Data for analysis and dictionaries of aggressiveness lexicon

2.1 Data from social media

The study was conducted on two types of materials from open Internet sources in Russian:

1) frequency distribution of words for December 2018. (the sum of all word is more than 32.6 billion); the low-frequency “tail” of the distribution was cut off at the level of 95% of the total number of unique words; as a result, the frequency dictionary contained at least 200 thousand unique words for each calendar day;

2) a small pool of Russian-language messages from social media, collected in October 2018 and marked by domains, gender and age (the total number of messages is more than 6.8 million).

Data was provided by monitoring and analysis system social media Brand Analytics.

The distribution in percent of Russian-language messages across domains is shown in Fig.1. The largest number of publications is instagram.com (29%), followed by vk.com (25%), facebook.com (20%), ok.ru (13%), twitter.com (6%) and youtube.com (2%). Social media distribution usually looks different on the full data stream, but the working series of data makes up about 1% of the total data stream, so there are some imbalances for several resources (for example, in the Russian-language sector, the palm of victory must belong to vk.com - see link [8]).

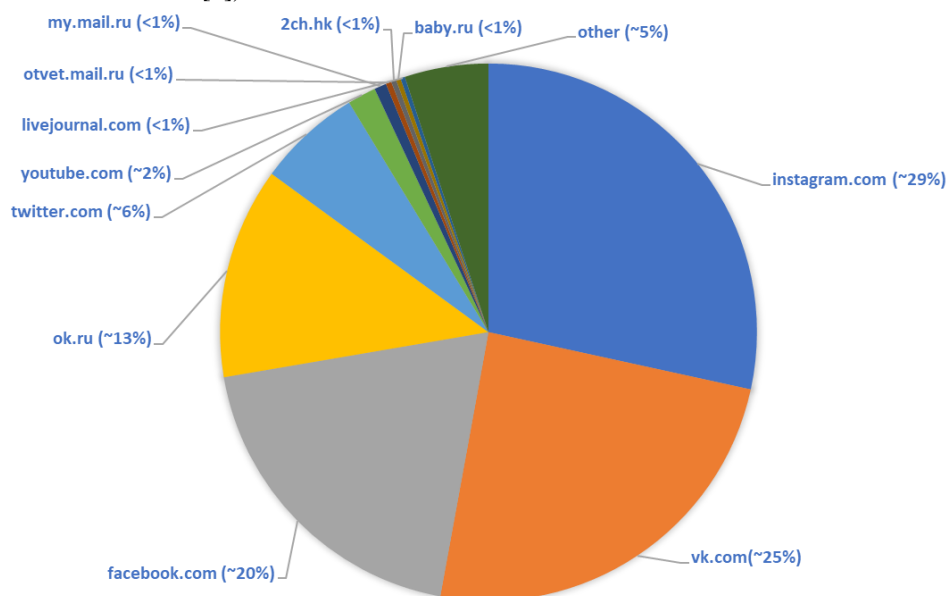


Fig.1. The first 11 domain distribution in per cent to the total number.

2.2 Dictionaries of aggressiveness lexicon

To study aggression and aggressiveness, our linguistic experts created dictionaries divided into parts of speech and types of aggression:

- dictionary of aggressive adjectives (qty ~ 200; for example, ‘быдловатый’, ‘дебильный’, ‘задрипанный’ etc.);
- dictionary of aggressive adverbs (qty ~ 100; for example, ‘эхидно’, ‘нефик’, ‘позорно’ etc.);
- dictionary of aggressive interjections (qty ~ 200; for example, ‘хренли’, ‘блиин’ etc.);
- dictionary of aggressive type A nouns: nominative invectives of a person (qty ~ 500; for example, ‘лох’, ‘нашистик’, ‘фуфел’ etc);
- dictionary of aggressive type B nouns: nominative invectives of an inanimate objects (qty ~ 90; for example, ‘адище’, ‘спербанк’, ‘рашка’ etc.);

- dictionary of aggressive type C nouns: nominative invectives as a reaction to an event (qty ~ 290; for example, 'баянище', 'мерзость', 'отстой' etc.);
- dictionary of aggressive verbs (qty ~ 470; for example, 'вбухивать', 'нафлудить', 'хренеть' etc.).

The total amount of aggressive vocabulary dictionaries is more than two thousand lemmatized words. The dictionaries include not only grammatically correct words, but also the most frequently used slang and misspelled words.

(«*фшоке*», «*чурко*», «*фошыст*» и т.д.).

In our research we didn't try to classify aggression into invective categories. Our study was an attempt to determine the functional types of aggressive discourse.

We tried to divide nouns according to the types of aggression described in section 1.2. Combinations of nouns with aggressive adjectives were also classified in texts: the word combination was determined by the animateness of the noun (provided that the noun is not marked by the type of aggression), the animated type was related to the type of aggression A, the inanimate to the type of aggression B.

Messages having aggression type D ("aggressive intentions or call for aggressive actions") contain very often verbs that are very difficult to identify unambiguously at the dictionary level («*послать*», «*засунуть*», «*лизать*» etc.). Therefore, this type of aggression, as well as its implicit form, was not yet analyzed.

All word forms were generated on the basis of these lemmatized lexicon taking into account the parts of speech. The result was more than 51 thousand word forms.

- adjectives - ~ 5900 word forms;
- adverbs and interjections ~ 215 word forms;
- nouns (type A) ~ 11,700 word forms;
- nouns (type B) ~ 2400 word forms;
- nouns (type C) ~ 9100 word forms;
- verbs - ~ 21900 word forms.

We have united interjections and adverbs because the functions of the invective adverbs and interjections often coincide, therefore their dictionaries overlap partially.

This volume was the basis for the studies of aggression and aggressiveness on the frequency distribution and a set of messages.

2.3 *Methods and criteria for the selection of vocabulary*

The dictionaries of aggressive vocabulary were based on the dictionaries of a system of object-oriented sentiment analysis [6] (total volume - more than 21 thousand words and phrases), from which words with high expressive weights were extracted. These dictionaries of negative, positive, as well as amplifying words vocabulary ((which can be used both in a negative and positive context, enhancing it, e.g. "ахренительный", "ваааааше", "упячка") were used.

A dictionary of aggressive sustainable phrases was also compiled, the words of which do not have an aggressive component themselves (for example, «баранки гнуть», «фильтровать базар», «лапшу на уши», etc.).

The marking conditions were:

- a predetermined type of aggression (for nouns);
- the weight of aggressiveness.

The words were divided into two groups. The first group included vocabulary, which almost always occurs in an aggressive context (for example, «быдло», «возбухать», «похрену») - they were marked with a unit weight. The second group included words that occur in an aggressive and non-aggressive context both (for example, «развалюха», «туповато», «гадостный») - they were marked a zero weight. This was done in order to take into account the weight of the word when calculating aggression.

Homonyms that have both aggressive and neutral meanings were not taken into account (for example, the imperative mood or the adverbial participle «юлить» coincides with some forms of the proper name «Юля», etc.).

The words, the aggressiveness of which depended on the context or discourse, were also not taken into account (for example, «козел» was not taken into account, but «казел» was, since such errors, as a rule, are made intentionally).

We rejected the statistical methods of weighting, because aggressive vocabulary is usually grass-roots level in frequency distribution, and this leads to strong errors in the statistical analysis of low-frequency components and depends on many factors of the analyzed data (subject, material type, collection time, etc.).

Marking was done by three linguists. The coincidence of the labels of least two annotators was a matching criterion. Some controversial cases were discussed separately.

2.4 Estimate of aggression and aggressiveness

A calculating of aggression and aggressiveness was based on:

- the total number of words in the message (N);
- type of the dictionary in which a word has an occurrence;
- the number of words in the message with zero weight (S_{null});
- the number of words in a message that have a unit weight (S_{one});
- normalization factor k.

The normalization factor k is needed to increase the weight of messages that have a number of words close to the average length of aggressive messages (so that very short messages do not get to the top).

The aggressiveness (A) of the message was determined when $S_{null} > 0$:

$$A = k * ((S_{null}-1)*p_0 + S_{one}) / (1+N - (S_{null} + S_{one})),$$

In other cases:

$$A = k * S_{one} / (1+N - S_{one})$$

Where p_0 (0.1-0.3) is an empirical coefficient influencing the total weight of aggressiveness.

The type of aggression was determined by the occurrence of the word in the appropriate dictionary.

Numerical thresholds were also introduced to filter out irrelevant data. Therefore, messages containing a small amount of aggressive vocabulary, zero-weight vocabulary were not marked as aggressive.

3 Results and Analysis

3.1 Frequency Distribution Analysis

About 5% of the 51 thousand word forms obtained were found in the frequency distribution.

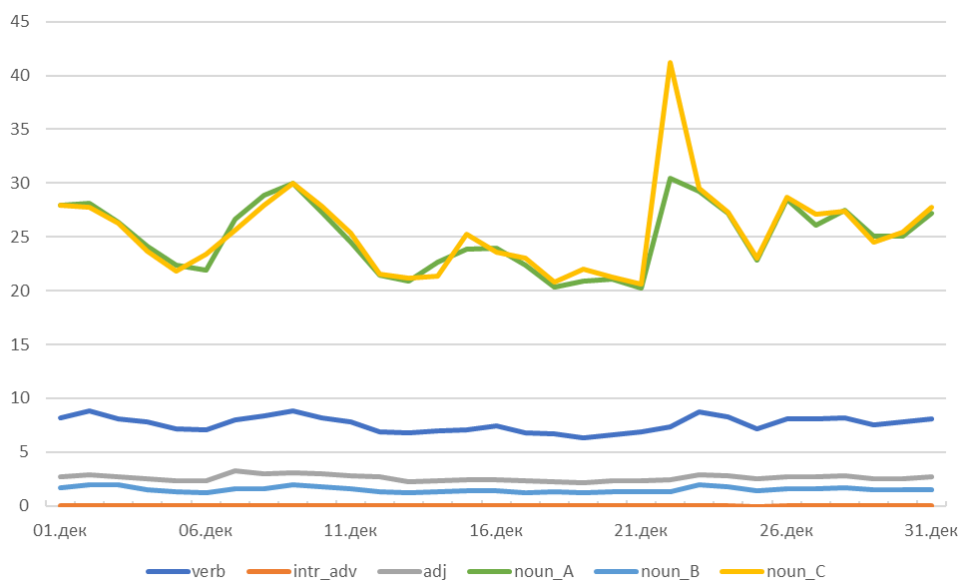


Fig.2. The distribution of aggressive vocabulary in parts of speech in absolute values for December 2018. Abbreviations: verb - verbs, intr_adv - interjections and adverbs, adj - adjectives, noun_A - nouns type A, noun_B - nouns type B, noun_C - nouns type C.

Figure 2 shows the distribution of aggressive vocabulary by parts of speech in absolute values (not normalized to the number of word forms). As it can be seen, the biggest part is occupied by type A and type C nouns; the third place belongs to the verbs, although the number of their word forms is twice bigger. It is possible to make an assumption about the predominance of the nominative function in aggression, in contrast to the sentiment analysis, which is predominantly predicative [6].

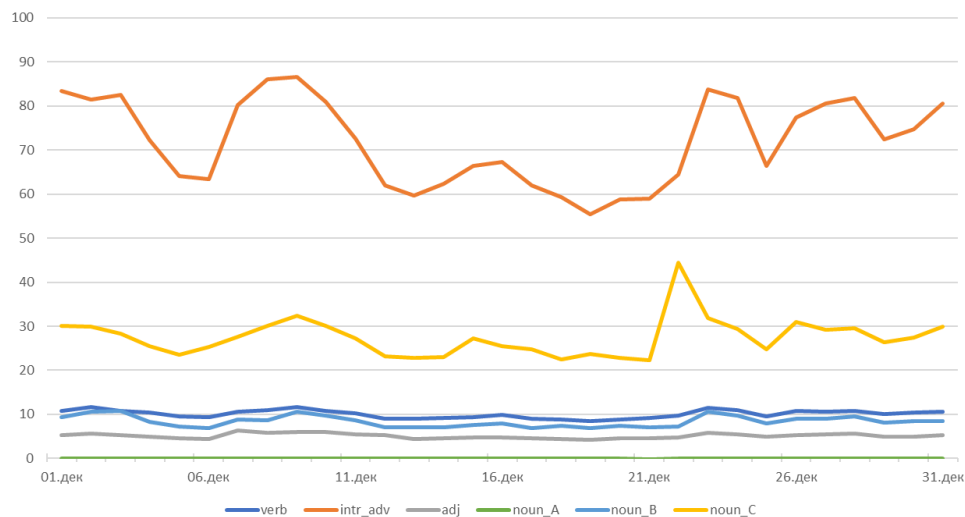


Fig.3. The normalized distribution of aggressive vocabulary in parts of speech in relative values.

Despite of the small volume of the dictionary, groups of adverbs and interjections show a big growth. Thus, mainly interjections and adverbs accumulate lexical aggressiveness (the intensity of use is about 70% of all parts of speech).

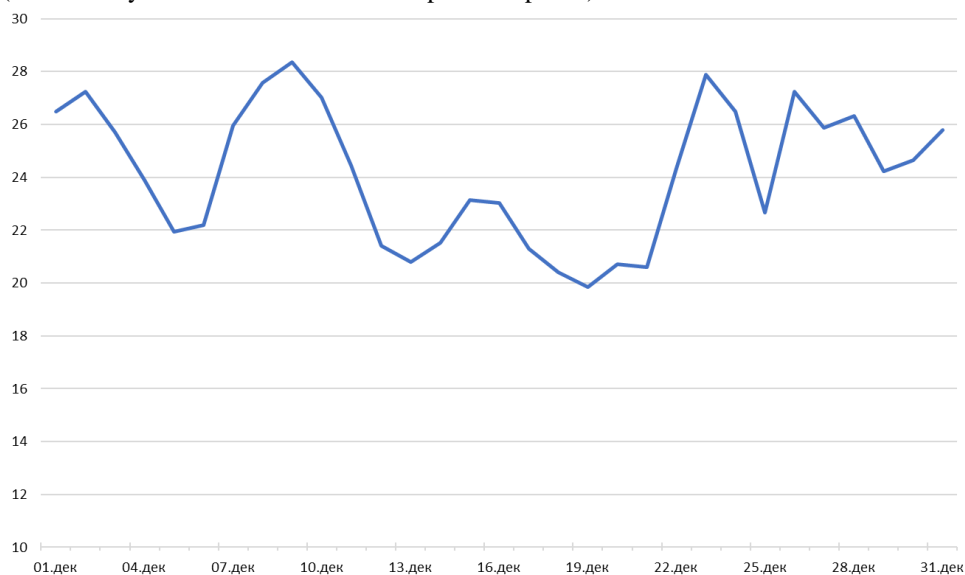


Fig.4. The total normalized distribution of aggressive vocabulary for December 2018.

Figure 4 shows the total normalized distribution of all aggressive vocabulary on the frequency distribution. As you can see, the shape of the parts of speech curves are similar. This means that aggressive vocabulary does not qualitatively depend on parts of speech, only quantitatively. On the frequency distribution, one can observe a general picture of the manifestation of aggressiveness. Aggression is a subset of aggressiveness, and to recognize its type, analysis of the message itself is necessary.

The correlation coefficient between the frequency distribution and aggressiveness in December 2018 is equal to 0.74, - this is a high enough similarity indicator. This is not surprising: in December 2018. there were no bright events, accompanied by aggressiveness. This result can be compared with December 2017, where the correlation coefficient was -0.79, which indicates its absence (see Fig. 5).

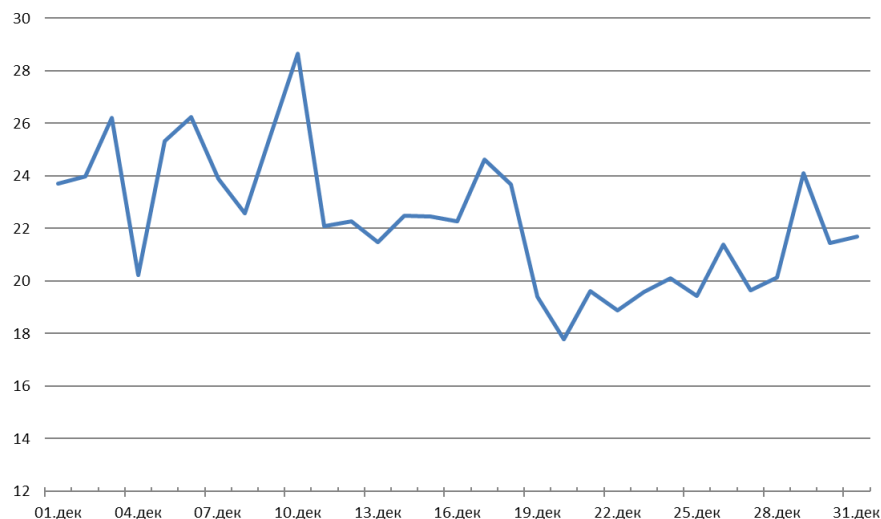


Fig.5. The total normalized distribution of aggressive vocabulary for December 2017.

In December 2017 several events occurred in the social and political life. For example, the growth of aggressiveness in early December was connected with exacerbation the situation around the Olympics, and the biggest peak was most likely caused by the IOC's punishing the Russian national team (the decision was made on December 5th, but the "ball" reached the 10th). The peak in mid-December seems to be related to the court decision on the minister Ulyukayev's case (December 15th), which was overlapped by the annual Press Conference of President V. Putin (December 14th). A similar peak (the annual Press Conference of President) of type C aggression is also observed in Figure 2 after December 20th.

Similar curves were obtained for some other months. For example, Fig. 6 shows the distribution of aggressive vocabulary for March 2018. The correlation coefficient for March between the frequency distribution and aggressiveness was relatively low (0.13).

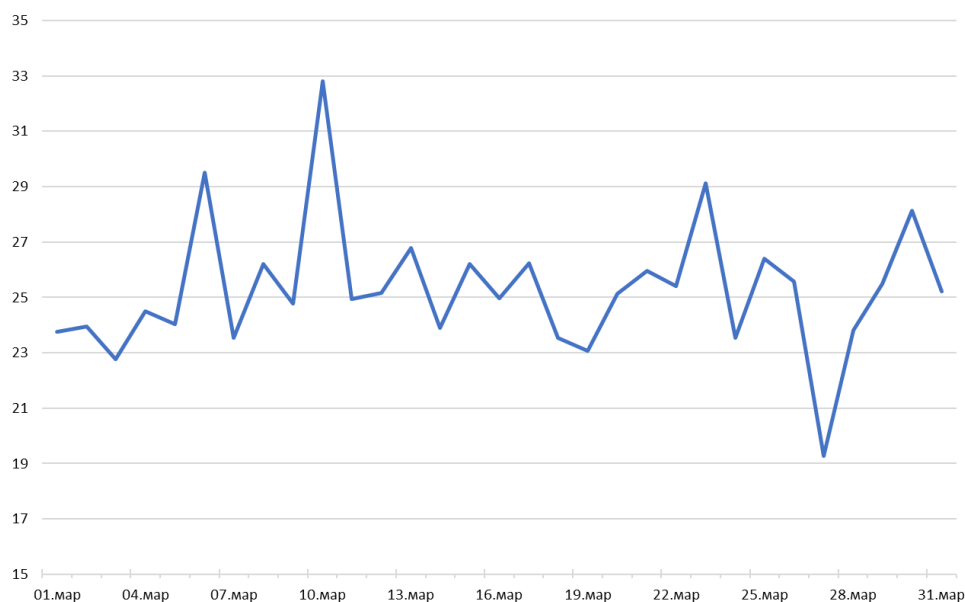


Fig.6. Total normalized distribution of aggressive vocabulary for March 2018

This month was also filled with landmark events of social and political life. A peak of 5-7th March corresponds to the first reaction to the Skripals poisoning (beginning on March 5th), and March 8-12th was the second reaction, when it became clear that everything is serious (active dissemination of news, expulsion of Russian diplomats from the UK). The peak around the 23rd was probably caused by the final court decision on blocking the Telegram messenger. And the cause of aggressiveness gap of March 25-27th is a fire in Kemerovo (March 25th). Finally, the rise of aggressiveness at the end of March is associated with the expulsion of Russian diplomats from the EU, the United States and Canada.

Analysis of social media messages

In total, more than 6.8 million messages were analyzed. Approximately 5% of the messages from all were aggressive.

3.1.1 Distribution by gender

More than half of the messages — 56% — have had a gender mark: more than 1.5 million males and more than 2.2 million females.

The relative aggressiveness of the male part is expected to be twice as high as that of the female: ~ 2.5% of the total of messages with a male mark, ~ 1.2% of the total of messages with a female mark.

The calculation of aggressiveness and aggression was based on an empirical metric that takes into account the length of the message, the type of aggression and the weight of aggressive vocabulary.

3.1.2 Age distribution

The graph of the distribution of aggressiveness by age is shown in Fig.7.

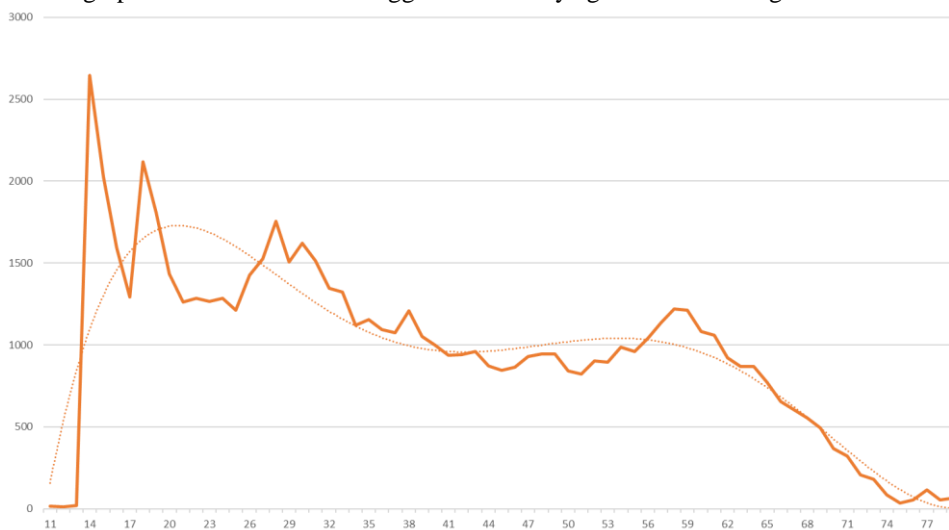


Fig.7. Distribution of aggressiveness by age. The abscissa is the age. The dotted line shows the polynomial trend line.

Figure 7 shows a peak around 14 years of age. This is most likely an artifact because most social networks allow registration from the age of 14 (for example, vk.com, which is the largest Russian-language social media) and the bots often set the default age.

3.1.3 Distribution by sex and age

The distribution of aggressiveness between the male and female part of society has different curves.

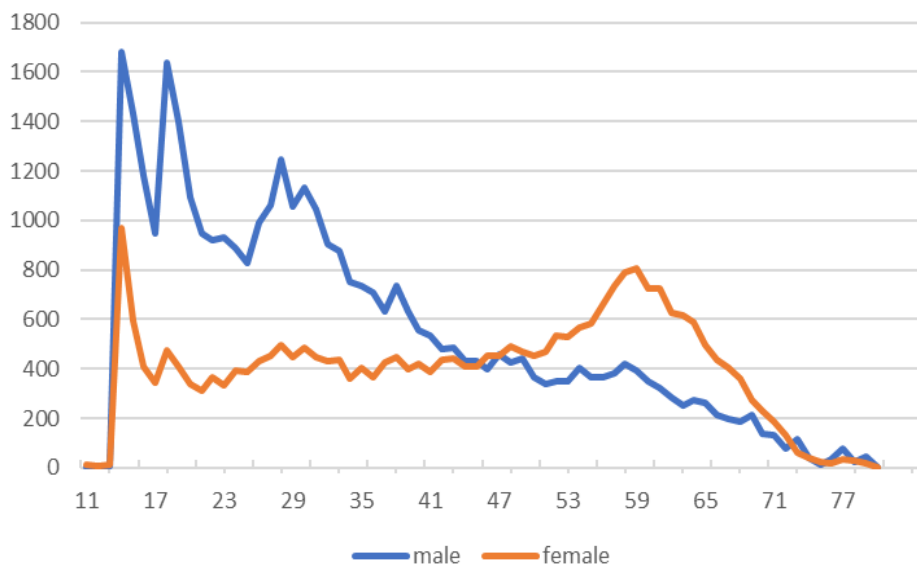


Fig.8. Distribution of aggressiveness by age and sex. The abscissa is the age.

Figure 8 shows the curves of the absolute distribution of aggressiveness between the male and female marks. Both curves have peaks in adolescence (14-19 years) and youth (25-35 years). Then in men, aggressiveness gradually decreases, and in women there is another peak at 55-65 years

3.1.4 Distribution across domain

The distribution of aggressiveness (fig.9) and aggression (fig.10) across domains showed that both are inherent, above all, to the entertainment domains of the young category (2ch.hk, yaplakal.com, pickabu.ru, etc.), then cybersport (prodota.ru, forums.goha.ru, dota2.ru, etc.), as well as forums where politics is discussed (echo.msk.ru, politforums.net, bolshoyforum.com, etc.). After that, women's sites (woman.ru, eva.ru) and car enthusiasts (forums.drom.ru, forum.auto.ru) go a little behind. It is noteworthy that the aggression rating of political platforms has relatively increased regarding their aggressiveness.

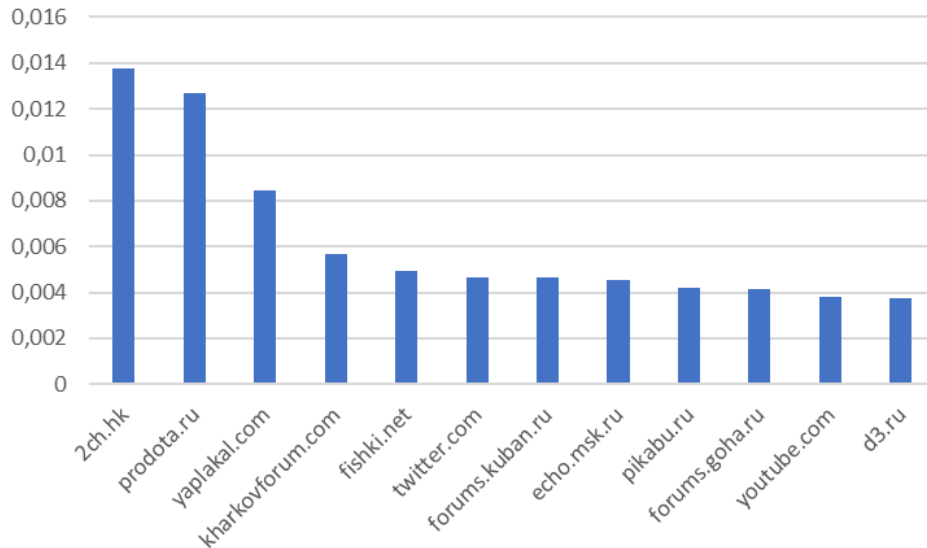


Fig.9. Distribution of aggressiveness in some Russian-language domains.

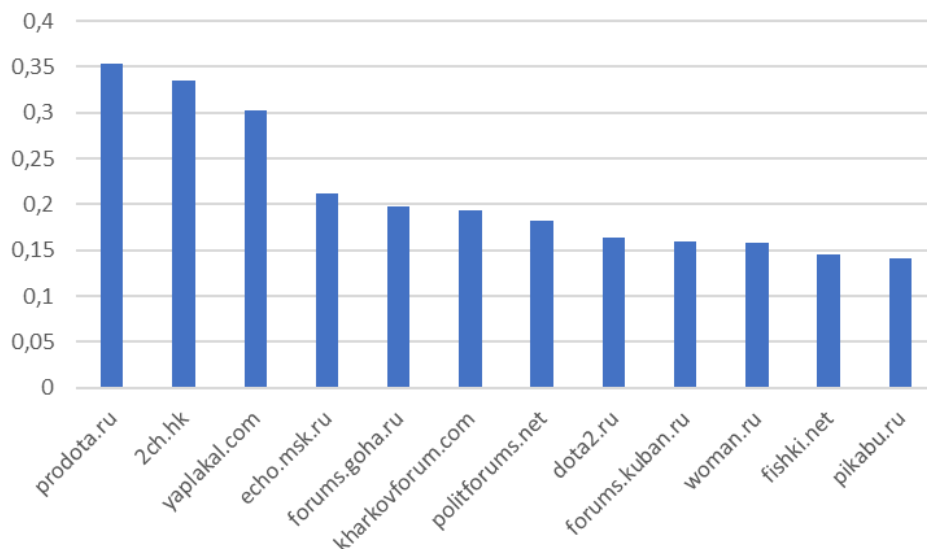


Fig.10. Distribution of aggression for some Russian-language domains.

At the same time, there is practically no correlation between the number of messages in domains with aggressiveness or aggression: -0.18 and -0.04, respectively.

3.1.5 Distribution by types of aggression

The analysis of the types of aggression in accordance with our classification showed that:

- the overwhelming number of messages contain a type C (reaction to an event) aggression - 86.2%;
- aggression directed at a person or a group of people (type A) - 13.9%;

- aggression directed at an inanimate object (type B) - about 2.1%;

A similar analysis of the types of aggression on a small Russian-language corpus showed similar results, with a slightly different ratio of types [9]:

- reaction to an event - 62.6%;
- aggression directed at a person or a group of people - 26.7%;
- aggression directed at an inanimate object - 2.8%;
- calls for aggression - about 8%.

The ratio of the total number of documents containing aggressiveness to the total number of documents containing aggression in this pull date was:

$$k = \frac{\textit{aggressiveness}}{\textit{aggressive}} = 0.86$$

4 Conclusions

The results of the analysis of the aggressive vocabulary does not depend qualitatively on parts of speech (only quantitatively).

The texts are dominated by aggressiveness, expressed mostly in adverbs and interjections.

Aggressive function as a type of behavior, expressed in different types of aggression, prevails in terms of volume and diversity in absolute meaning (Fig. 2), but inferior in frequency of use of the cathartic function, manifested mainly in adverbs and interjections (Fig. 3).

Age analysis showed that the curves of female and male aggressiveness differ (Fig. 8): female aggressiveness has an additional rise at the age of 55-65 years.

Analysis of the type of aggression in social media showed predominance of type C, i.e. caused by reaction to events. However, this value may vary depending on the emotional nature of the events themselves.

The distribution of aggression and aggressiveness across domains revealed three main categories: youth entertainment communities, game portals and sociopolitical platforms.

The purpose of ours work is to study aggressive vocabulary on real materials of social media, an attempt to classify it not by invective categories, but by types of functional use. The study is in initial stages. Unfortunately, we did not have time to include in the presented work some interesting material on the classification of aggressive verbs, as well as identifying the type of "calls for aggression and aggressive actions." Consideration of context and discourse, as well as the definition of implicit aggression is under development.

In the future, this system of identifying aggressiveness and determining the type of aggression is supposed to be used in conjunction with the system for automatically determining trends in media space. In this case, it will be possible to almost online determine the response to an event, to identify hot spots, the direction of aggression.

Summing up, it is safe to say that aggressiveness and aggression are important markers of the social activities of people and require further study.

Reference

- 1.ACL 2017. The First Workshop on Abusive Language Online. Proceedings of the Workshop. Available at:<http://aclweb.org/anthology/W17-30>
- 2.Apresjan V.Ju. (2003) Implicite aggression in language [Implizitnaja argessija v jazyke] Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2003" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii "Dialog 2003"], pp. 32-35.
- 3.COLING 2018. First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018). Proceedings of the Workshop. Available at: <http://aclweb.org/anthology/W18-44>
- 4.Gray W. (1973) Emotional-Cognitive Structuring: A New Theory of Mind // Forum of Correspondence and Contact. – 1973.V5 – P.1-6.
- 5.Izard C. Human Emotions. (1977). New York: Plenum.
- 6.Pazelskaya A., Soloviev A. (2011) A method of sentiment analysis in Russian texts [Metod opredelenia emozij na russkom yazyke] Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2011" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii "Dialog 2011"], Vol. 11(18), pp.510-523.
- 7.Pinker, S. (2007). The Stuff of Thought : Language as a Window Into Human Nature [Substanzija myshlenija. Jazyk rar okno v chelovecheskuju prirodu], New York, NY: Viking.
- 8.Social Media in Russia: Digits and Trends, Autumn 2018. Available at: <https://br-analytics.ru/blog/socseti-v-rossii-osen-2018/>

9. Soloviev A.N. (2018) Aggressiveness and aggression on the materials of social media, A new culture of communication in the context of digital and sociocultural globalization: law, media and national identity, International Scientific and Practical Conference, Moscow.
10. Waseem Z., Davidson T., Warmusley D. and Weber I. (2017), Understanding Abuse: A Typology of Abusive Language Detection Subtasks / Proceedings of the First Workshop on Abusive Language Online, pp. 78–84.
11. Комалова Л. Р. Язык и речевая агрессия (аналитический обзор). М.: ИНИОН РАН, 2015.
12. Пилецкий С.Г. Каиново наследство, или Опыт изучения ЧЕЛОВЕЧЕСКОЙ АГРЕССИВНОСТИ: Анализ этологических, социобиологических и современных социально-гуманитарных концепций. М.: URSS. 2015.
13. Шаховский В.И. Категоризация эмоций в лексико-семантической системе языка. М.: Гнозис, 2008.